

Quels pays pour exporter notre production de poulet ?

CLADIERE Nathan, projet 5, Formation Data Analyst OC

Sommaire

- ▶ Contextes et objectifs
- ▶ Préparations des données
- ▶ Première étude
 - ▶ Clustering hiérarchique
 - ▶ Différences significatives entre clusters
 - ▶ Visualisation et caractéristiques des clusters par ACP
- ▶ Deuxième étude, données supplémentaires
 - ▶ Clustering par K-means
- ▶ Conclusion, pays retenus

Contexte/objectifs

- ▶ Première étape d'une étude de marché:
 - ▶ Exportation de poulet à l'international
- ▶ Utilisations de données accessibles facilement
 - ▶ FAO
 - ▶ Banque mondiale
- ▶ Etudes des caractéristiques des différents pays
- ▶ Trouver des pays cible pour approfondir l'étude de marché

Contexte/objectifs: données

- ▶ Source FAO:
 - ▶ Population: entre 2014 et 2017
 - ▶ Protéines animales disponibles dans le pays
 - ▶ Protéines totales disponibles dans le pays
 - ▶ Disponibilité alimentaire:
 - ▶ Protéines
 - ▶ Calories
 - ▶ Habitudes avec la volailles:
 - ▶ Import
 - ▶ Export
 - ▶ Nourriture
 - ▶ Production
- ▶ Source Banque mondiale:
 - ▶ PIB par habitant 2017

Préparation des données

- ▶ Données manquantes:
 - ▶ Population, Pib 2017: recherche et implémentation
 - ▶ Détails volailles : remplacées par 0 (faible ou peu mesurable)
- ▶ Calcul de la croissance de population 2014 - 2017
- ▶ Calcul du ratio protéines animales/protéines totales disponibles

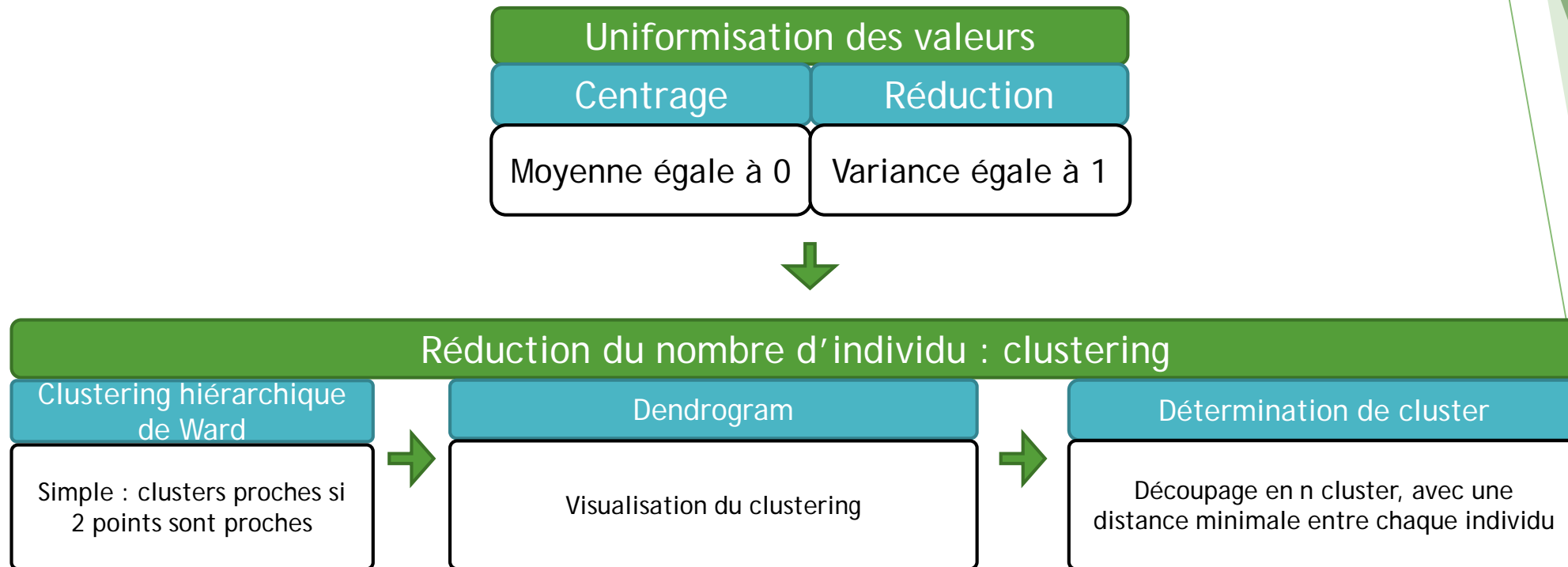
Première étude

Première étude: Données étudiées

- ▶ 172 pays
- ▶ 4 Variables:
 - ▶ Croissance de la population de 2014 à 2017
 - ▶ Disponibilité alimentaire
 - ▶ Protéines
 - ▶ Kcal
 - ▶ Ratio protéines animales disponibles / protéines totales disponibles

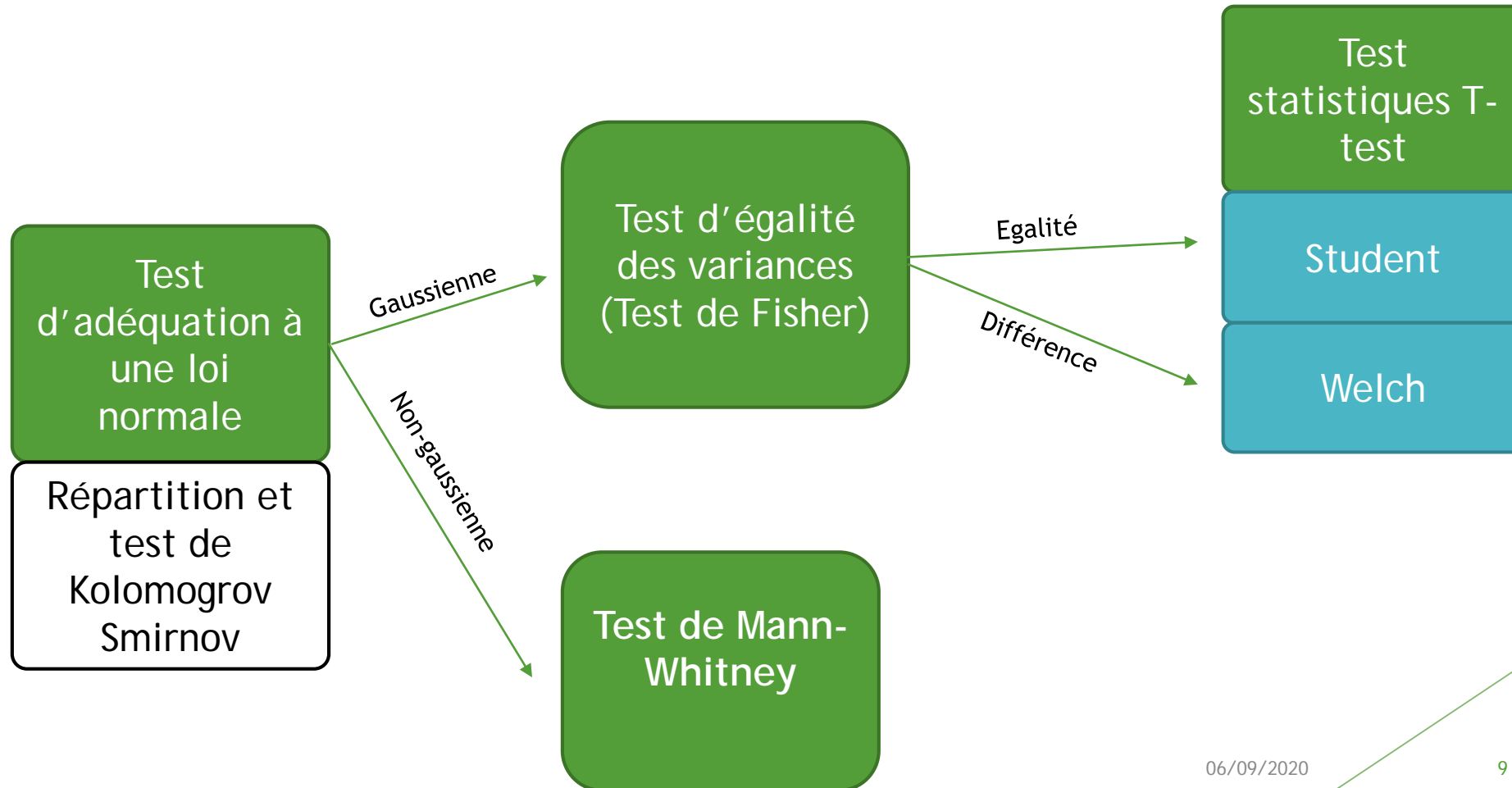
Première étude: démarches

Uniformisation des valeurs et Clustering



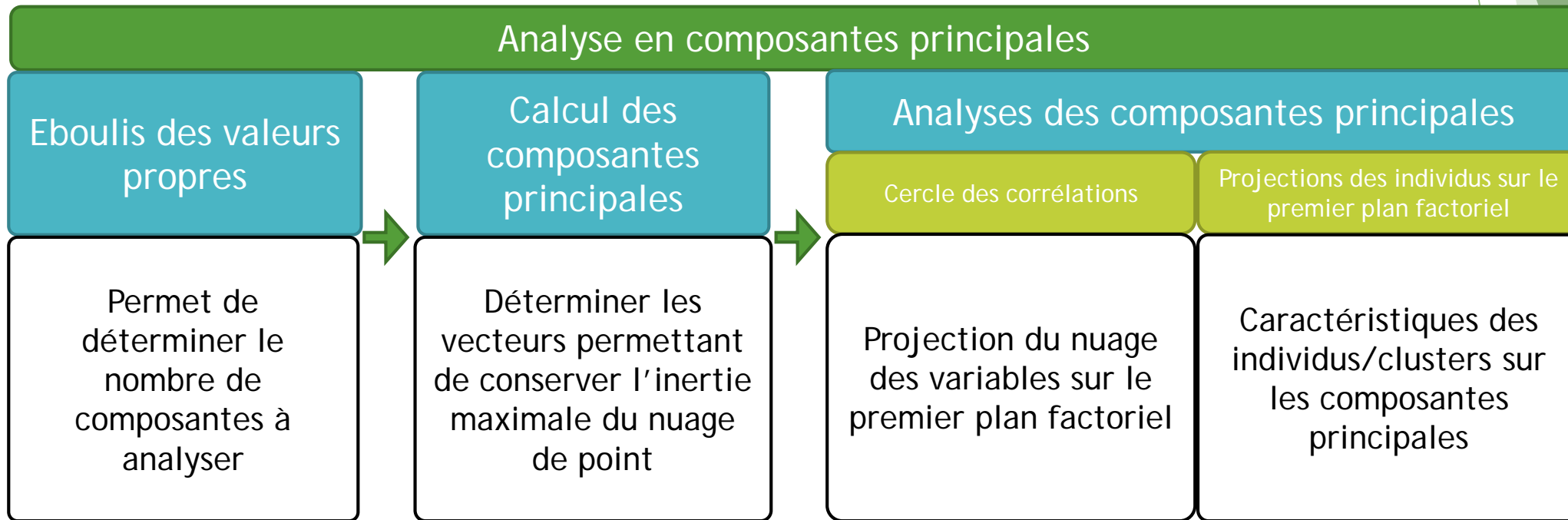
Première étude: démarches

Différences entre clusters



Première étude: démarches

Réduction du nombre de variables



Première étude, clustering hiérarchique: Centrage et réduction

- ▶ Centrage et réduction des données

- ▶ Exemple $x_{cr} = \frac{x - \bar{x}}{s_x}$

- ▶ Disponibilité en protéine Afghanistan $_{CR} = \frac{19,7 - 29,3}{7,3} = -1,31$

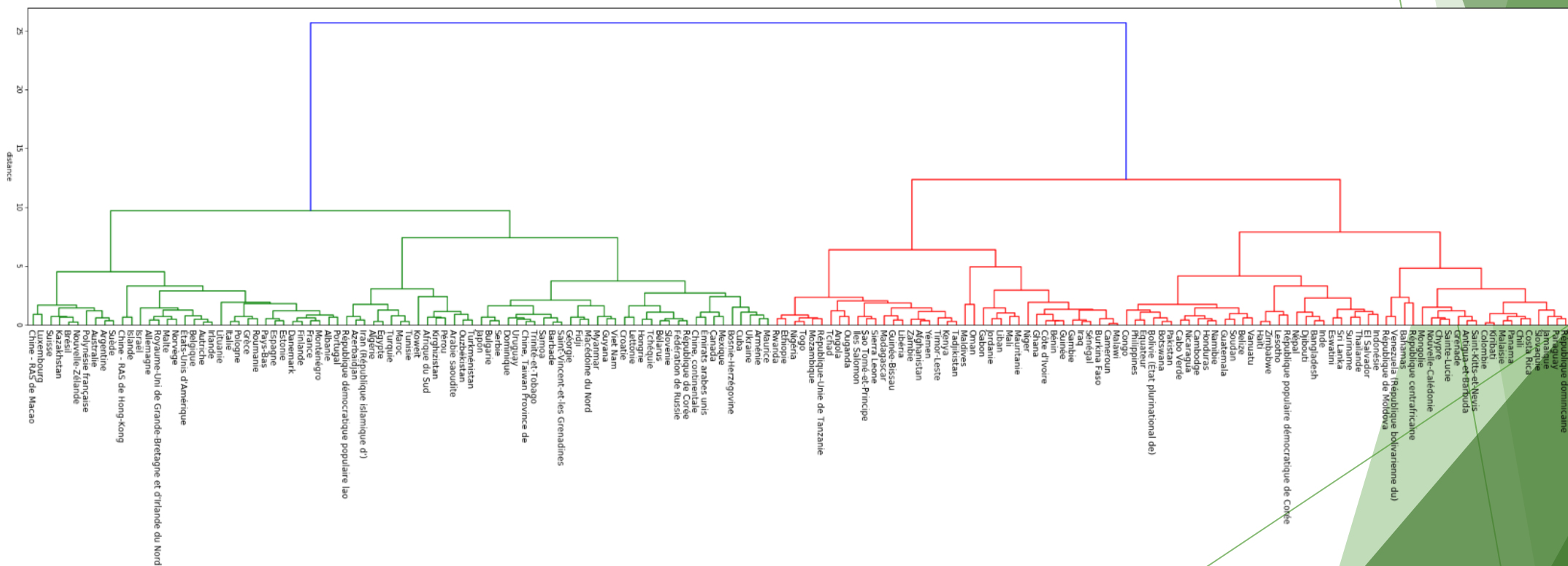
- ▶ Sélection des données `X = dataClustering.values`

- ▶ Centrage `std_scale = preprocessing.StandardScaler().fit(X)`

- ▶ Réduction `X_scaled = std_scale.transform(X)`

Première étude, clustering hiérarchique: Dendrogramme

► Clustering hiérarchique de ward



Première étude: Clustering

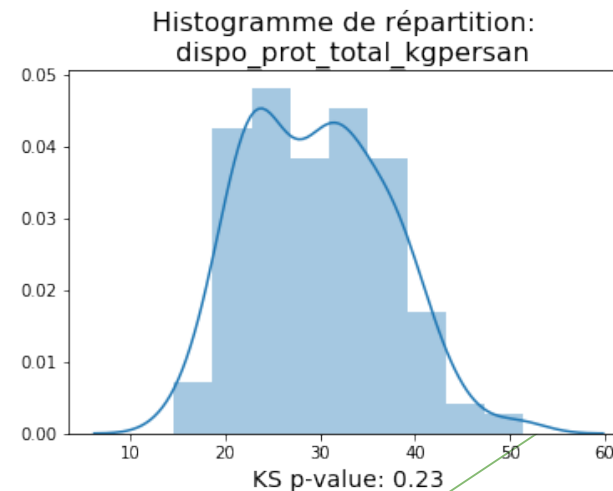
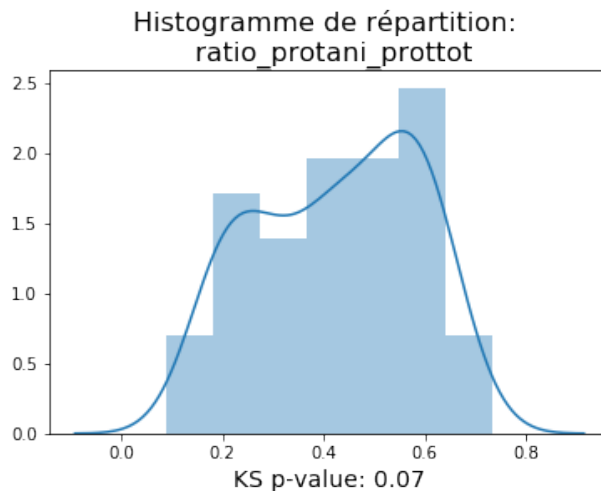
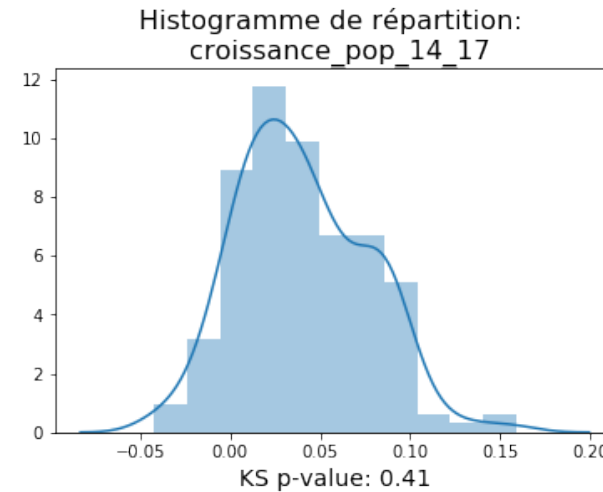
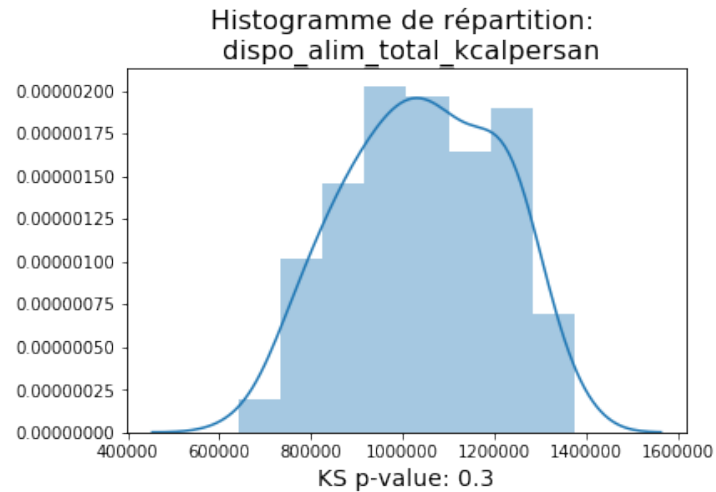
- Découpage du clustering de Ward en 5 clusters :

```
clusters = fcluster(Z, 5, criterion='maxclust')
```

Clusters	1	2	3	4	5
Nombre de Pays	35	48	39	29	21

- Première approche
 - Groupe de pays vraiment différents ?
 - Caractéristiques de ces groupes ?
 - Premier clustering suffisant pour le choix ?

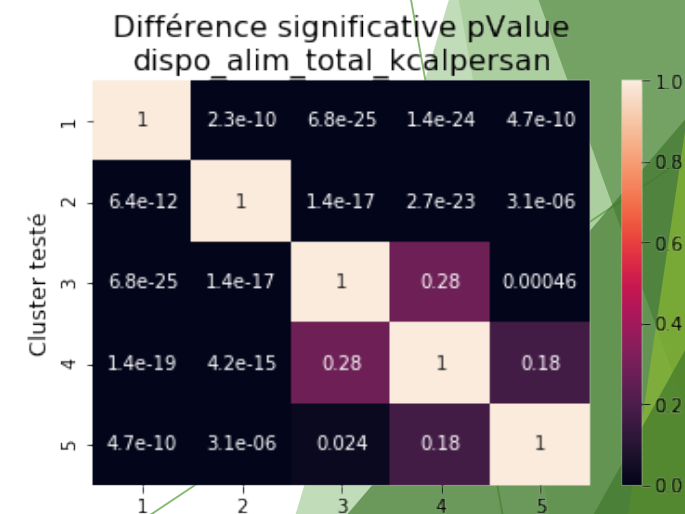
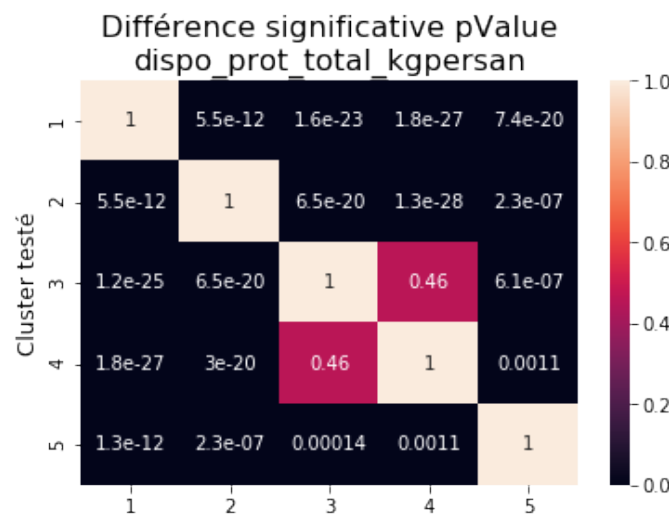
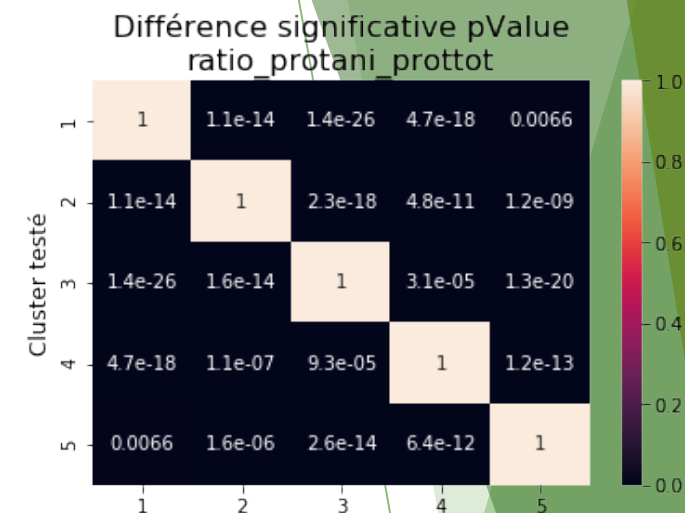
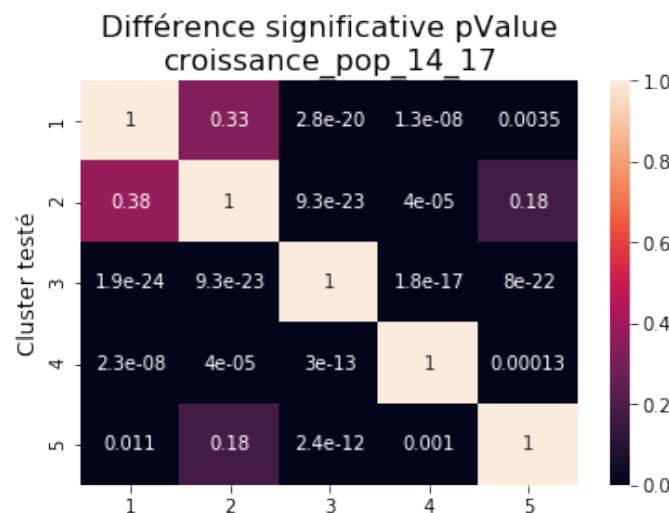
Première étude: test d'adéquation à une loi normale



Première étude: Test de différence significative entre clusters

Similarités entre clusters:

- Croissance de population:
 - 1 et 2
 - 2 et 5
- Disponibilité alimentaire
 - Protéine
 - 3 et 4
 - Kcal
 - 3 et 4
 - 4 et 5

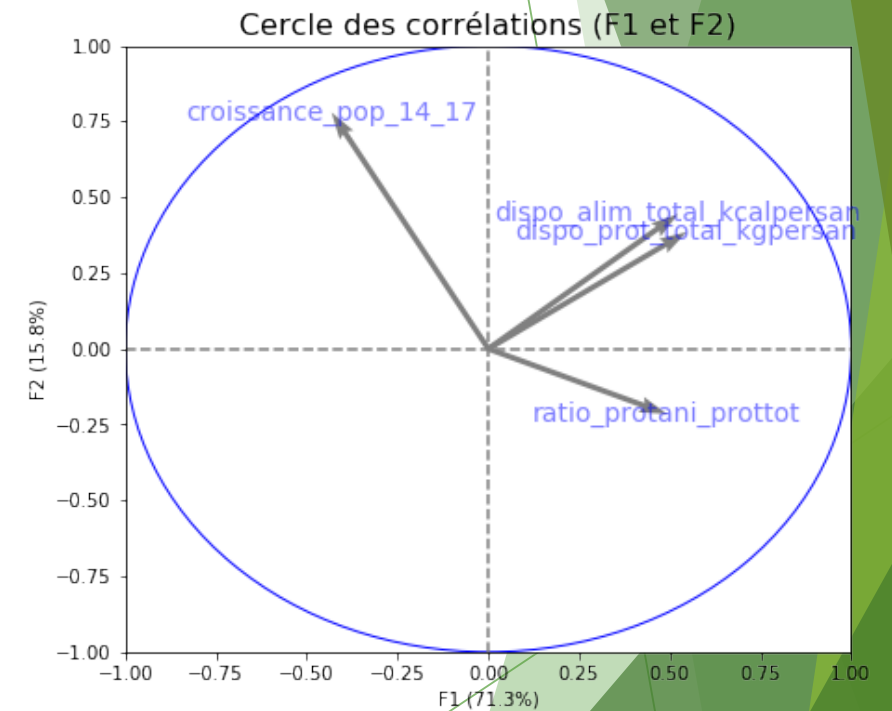


Première étude:

Calcul des composantes principales

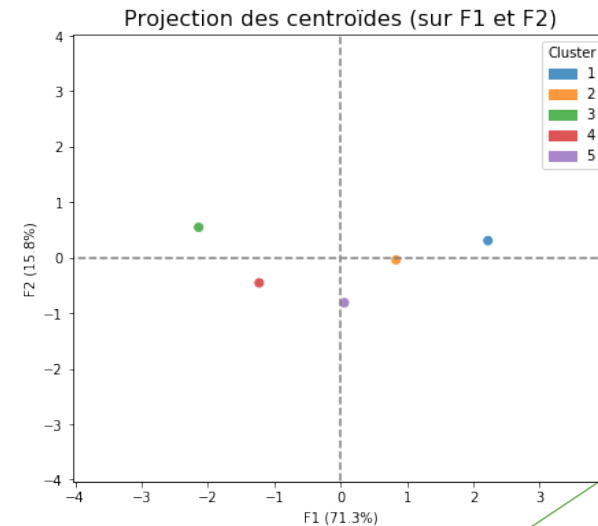
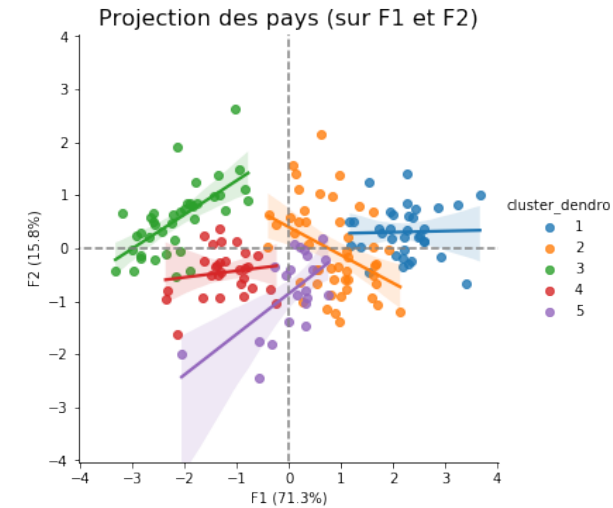
Variables Composantes	croissance_p op_14_17	ratio_protani_ prottot	dispo_prot_to tal_kgpersan	dispo_alim_tota l_kcalpersan
F1	-0.43	0.49	0.55	0.52
F2	0.78	-0.21	0.39	0.44

- F1 : disponibilité alimentaire, ratio protéines animales
- F2: Croissance de population



Première étude: Projection des centroïdes sur le premier plan factorielle

- Cluster 1 peu intéressant:
 - Pays avec une forte disponibilité alimentaire (39 kg)
 - Faible croissance de population
- Croissance de population : plus de clients potentiels
 - Le cluster 3 semble plus intéressant
- Pays avec une alimentation basée sur la viande :
 - Clusters 1 et 5
- Pays ayant besoin de viande
 - Clusters 2, 3, 4 et 5 (32 kg ou moins en moyenne)



Première conclusion

- ▶ Pays à ne pas cibler
 - ▶ Suffisance en protéines et Kcal
 - ▶ Faible croissance population
- ▶ Besoins de nouvelles données pour cibler plus précisément:
 - ▶ Capacité des personnes à acheter nos produits
 - ▶ Habitudes des pays avec les viandes de volailles

Deuxième étude

Deuxième étude : Nouvelles données

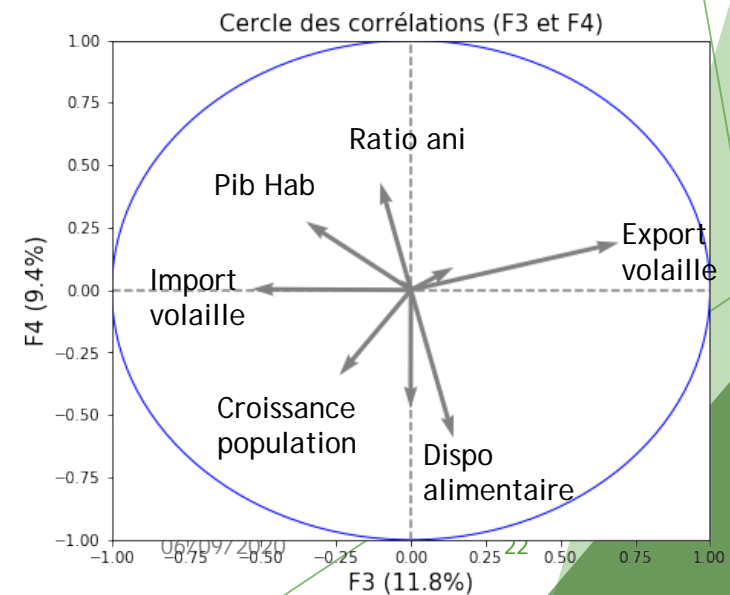
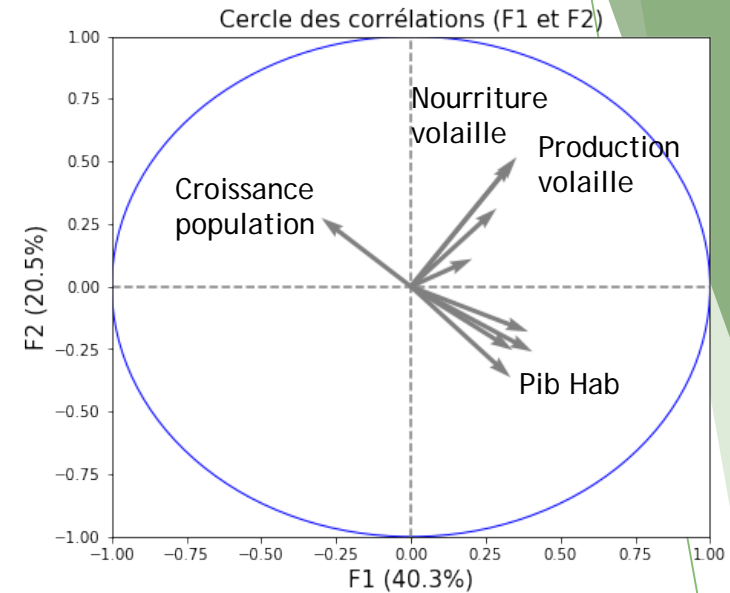
- ▶ Capacité des gens à acheter : PIB par habitant (banque mondiale)
- ▶ Habitudes avec la volaille (FAO)
 - ▶ Importation
 - ▶ Exportation
 - ▶ Nourriture
- ▶ Suppression du cluster 1

Deuxième étude : démarches

- ▶ Réduction du nombre variable: Analyse en composantes principales
 - ▶ Détermination des composantes principales
 - ▶ Interprétation du cercle de corrélations
- ▶ Clusterisation par K-means
 - ▶ Détermination du nombre de cluster par la méthode du coude
- ▶ Interprétation de la projection des centroïdes sur les composantes principales
- ▶ Choix des clusters et des cinq pays à étudier

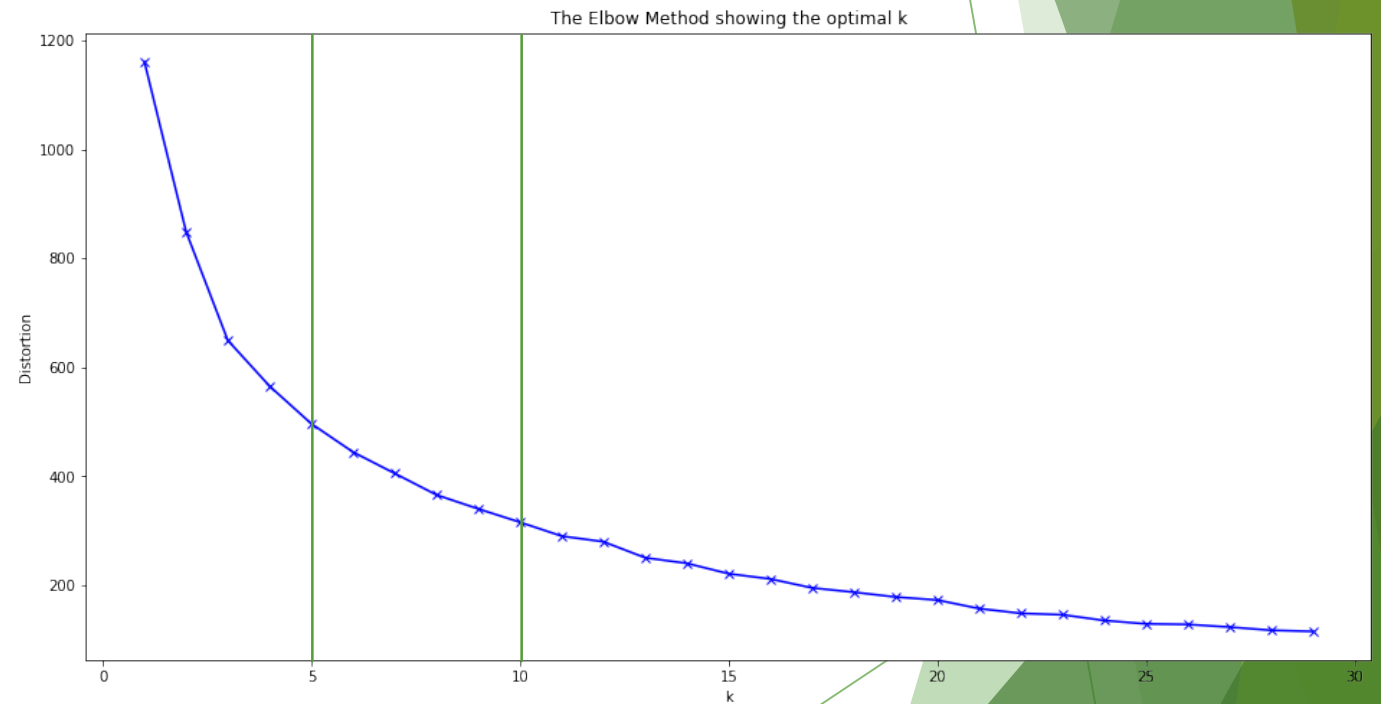
Deuxième étude: Calcul des composantes principales

- F1:
 - Positivement : Disponibilité alimentaire, pib Habitant
 - Négativement : croissance de la population
- F2: nourriture et production de volaille
- F3:
 - Positivement : export de volaille
 - Négativement : import de volaille
- F4 : Redondance avec F1



Deuxième étude: Clusterisation par K-means

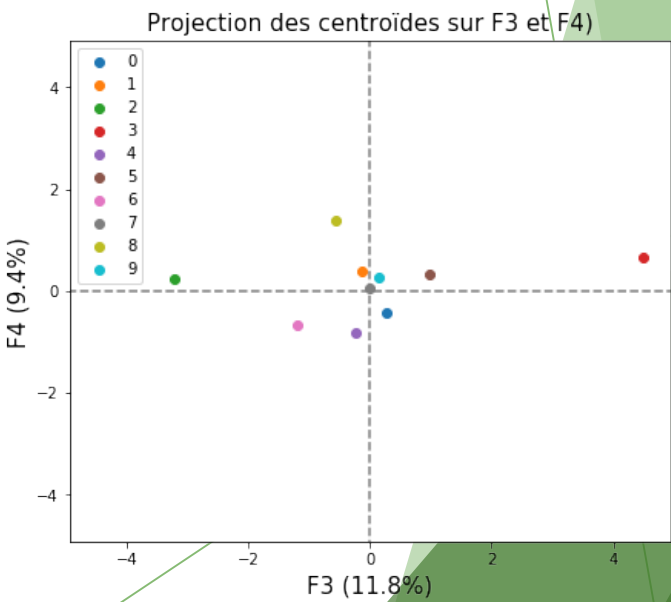
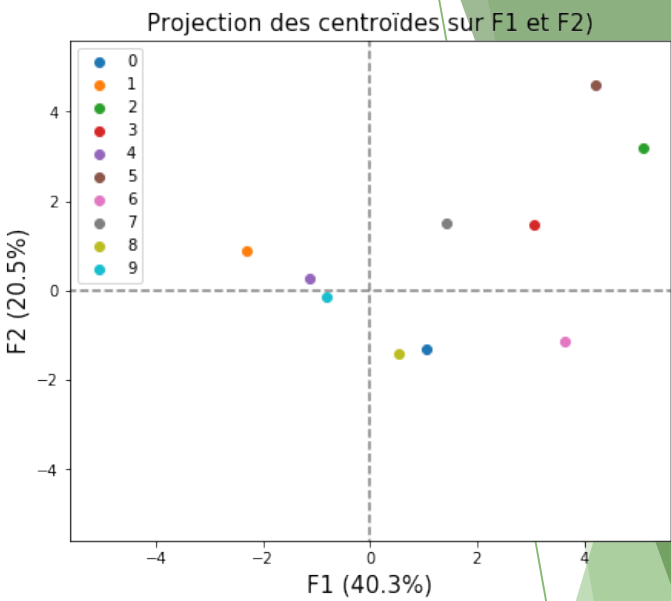
- ▶ Aplatissement à partir de 5 clusters
- ▶ Plus marqué encore à 10 clusters
- ▶ Test à 6 clusters : clusters trop volumineux
- ▶ Test à 10 clusters: meilleurs répartition dans les clusters



Choix du cluster avec les nouvelles données

Composantes					F1			F2	F3	
Clusters					PIB par Habitant	Croissance de population	Disponibilité alimentaire	Consommation et production de volaille	Importation de volaille	Exportation de volaille
2					++	--	--	++	++	++
6					++	-	--	-	+	+
	7				+	-	-	+	=	=
		1			--	++	++	=	=	=
		4			-	++	+	=	=	=
		9			-	+	+	=	=	=
			5		+	--	--	++	+	-
			3		+	--	--	+	--	--
				0	+	-	-	-	=	=
				8	++	-	-	--	=	=

+ et vert: intéressant pour l'exportation de nos produits
 - et rouge: frein à l'exportation de nos produits
 = et orange: pas de valeurs notables



Deuxième étude: Vérification de la différence significative

Clusters ciblés : 2 et 6

- ▶ Capacité à acheter : 22K\$ et 38K\$ en moyenne ($p=0,05$)
- ▶ Nombre de consommateurs potentiels :
 - ▶ pas de différence significatives avec les autres clusters
 - ▶ En moyenne assez basse (0,03) face aux pays en forte croissance: (0,09)
- ▶ Importation de volailles
 - ▶ 2 plus gros importateurs de volailles (900K, et 280K tonnes)
- ▶ Consommation de volailles
 - ▶ $p>0,05$ avec les gros exportateurs de volailles (5 et 3)

Clusters écartés

- ▶ 1,4 et 9 : pas en capacité d'acheter nos produits moins de 4k\$
 - ▶ Cluster 1: significativement différent de tous les clusters
- ▶ 5 et 3 : concurrents, gros exportateurs de volailles
 - ▶ Plus marqué avec le 3 ($p>0,05$ uniquement avec le 5)
- ▶ 8 et 0 : pas de caractéristiques significativement intéressantes
- ▶ 7 : pistes supplémentaires

Deuxième conclusion: Sur quels pays doit-on approfondir nos recherches

